

Query Match	92.7%; Score 257.8; DB 10; Length 279;
-------------	--



```

/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/Note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The U1-R-10
library is a subtracted library derived from an
individually-tagged normalized whole-eye (minus the lens)
library. The driver for the subtraction consisted of a
pool of all previous libraries (U1-R-A0, U1-R-A1, U1-R-E0,
U1-R-E1, U1-R-C0, and U1-R-C1). The tag is a string of
3-5 nucleotides present between the Not I site and the
oligo-dT track which allows identification of the library
of origin of a clone within the mixture. The subtracted
library (U1-R-Y0) was constructed as follows: PCR
amplified cDNA inserts from previous library clones from
which 3' ESTs had been derived were used as a driver in a
hybridization with the normalized whole-eye library in
the form of single-stranded circles. The remaining
single-stranded circles (subtracted library) was purified
by hydroxapatite column chromatography, converted to
double-stranded circles and electroporated into DH10B
bacteria (Life Technologies) to generate the U1-R-Y0
library. This procedure has been previously described
(Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
1996)."
BASE COUNT      123 a      112 c      105 g      76 t
ORIGIN

```

```

Query Match      73.0%; Score 203; DB 10; Length 416;
Best Local Similarity 84.8%; Pred. No. 2.9e-40;
Matches 224; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

```

```

QY 1 gccaaacccantccagacagtcgacgcatgcatgagatgcncaagtcac 60
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 152 gccAACACGCGCTGACGACGCGCGGCGATTGCCAATGCTCAAGTTAAC 211
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 gagcacatcaccaagctaaagctcagtcacattcatacagggaaggagctggcc 120
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 212 GACCAATTAACGCGCTTAACGCGAGTGCACTTCATCACCGGAGGAGATCTGGCC 271
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 atcatgagagctctccagacacaaagctcagagagctcggtttcatacagagag 180
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 272 ATCATGAGAGCGCTCGACGACACACAGTCTCACCGAGCTCGCTCCACACAGAGG 331
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 cacatcatgagcagcagctgagaaatgtaagatgtcaagctctcgaaggagaacagagc 240
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 332 CACATCATGGGCTCCAGATCGAGATGGAGATTGTCAACTGCTCAAGAGATATACACAG 391
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 ctctcagagctgggntaccattt 264
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 392 CTGCTGAGGCTGGGCTACCATTTT 415
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 4
W12756 468 bp mRNA linear EST 02-OCT-1997
LOCUS ma68a06.r1 Soares mouse p3NNMF19.5 Mus musculus cDNA clone
DEFINITION IMAGE:315826.5; similar to gb:54162_rnal 64 KD AUTOANTIGEN D1
(HUMAN); mRNA sequence.
ACCESSION W12756
VERSION W12756.1 GI:1286794
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

```

```

REFERENCE
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (Bases 1 to 468)
Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Weising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

```

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)  
CONTACT: Maria M/Mouse EST Project  
COMMENT WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI: 206442  
Seq primer: mob.REGA+ET  
High quality sequence stop: 403.  
Location/Qualifiers  
1. 468  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:315826"  
/clone\_lib="Soares mouse p3NNMF19.5"  
/dev\_stage="19.5 dpc total fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/Note="Vector: pT73D (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was primed with a Not I-oligo(dT) primer [5'  
TGTTACCAATCTGAACTGGAGCGCGGCGCATTTTCTTTTCTTTT 3']  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Benito  
Soares and M.Fatima Bonaldo. RNA was kindly provided by  
Dr. Minoru Ko (Wayne State University)."

BASE COUNT 144 a 118 c 120 g 85 t 1 others  
ORIGIN

```

Query Match      69.4%; Score 193; DB 10; Length 468;
Best Local Similarity 84.6%; Pred. No. 8.9e-38;
Matches 236; Conservative 0; Mismatches 41; Indels 2; Gaps 2;

```

```

QY 1 gccaaacgacantccagacagtcgacgcatgcatgagatgcncaagtcac 60
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 97 GCCAACACGCGCTGATGAGAGCGCGC -CGCATTCGCAATGACAGACATCAAGTTAAC 155
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 gagcacatcaccaagctaaagctcagtcacattcatacagggaaggagatctggcc 120
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 156 GACCAATTAACGCGCTTAACGCGAGTGCACTTCATCACCGGAGGAGATCTGGCC 215
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 atcatgagagctctccagacacaaagctcagagagctcggtttcatacagagag 180
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 216 ATCATGAGAGCGCTCGACGACACACGCTGCTCACGGAGCTCGCTCCACACACAGCGG 275
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 cacatcatgagcagcagctgagaaatgtaagatgtcaagctctcgaaggagaacagagc 240
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 276 CACATCATGGGCTCTCAGATGGAGATGGAAATTGTCAAGCTGCTCAAGAGATATACACAG 335
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 ctctcagagctgggntaccatt-ttnaactccaggacc 278
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 336 CTGCTGAGGCTGGGCTACCATTTTGAAGTCCACAGAGACC 374
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 5
A2985110 459 bp DNA linear GSS 27-APR-2001
LOCUS A2985110/c
DEFINITION 2M0266C19R Mouse 10kb plasmid U06C2M library Mus musculus genomic
clone U06C2M0266C19 R, DNA sequence.
ACCESSION A2985110
VERSION A2985110.1 GI:13856337
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

```

```

REFERENCE
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (Bases 1 to 459)
Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Weising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

```





CC and proteins involved in gro

The invention relates to polynucleotides (I) encoding diagnostic and therapeutic (DITHP) polypeptides (II), which include e.g. enzymes, and proteins involved in growth and development and receptors. (I) and

OM nucleic - nucleic search, using sw model

Run on: June 15, 2002, 09:52:18 ; Search time 69.62 seconds  
(without alignments)  
980.840 Million cell updates/sec

```

Title:          US-09-445-362B-6
Perfect score:  278
Sequence:       1  gccacacgcantcgcagca.....catttnactcccgacc 278

```

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Total number of hits satisfying chosen parameters: 7670666

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```
Database : Issued_Patents_NA:*
1: /cg2_6/prodata/1/ina/5A.COMB.seq:*
2: /cg2_6/prodata/1/ina/5B.COMB.seq:*
3: /cg2_6/prodata/1/ina/6A.COMB.seq:*
4: /cg2_6/prodata/1/ina/6B.COMB.seq:*
5: /cg2_6/prodata/1/ina/PCrUS.COMB.seq:*
6: /cg2_6/prodata/1/ina/backfiles1.seq:*
```

## SUMMARIES

Result	No.	Score	Match	Length	DB	ID	Description
C	1	39.4	14.2	52297	4	US-09-426-436-1	Sequence 1, Appl1
C	2	39.4	14.2	52297	4	US-08-705-557-1	Sequence 1, Appl1
C	3	35.2	12.7	538	4	US-09-183-959-13	Sequence 12, Appl1
4	4	34.8	12.5	687	5	PCT-US94-04361-25	Sequence 23, Appl1
5	5	34.6	12.4	681	5	PCT-US94-04361-27	Sequence 22, Appl1
6	6	33.8	12.2	2370	2	US-08-838-2119B-19	Sequence 17, Appl1
7	7	33.8	12.2	2370	3	US-09-233-536A-19	Sequence 19, Appl1
8	8	33.8	12.2	2370	3	US-09-233-552A-19	Sequence 19, Appl1
9	9	33.8	12.2	2370	4	US-09-402-036-19	Sequence 19, Appl1
10	10	33.8	12.2	2403	1	US-08-471-033-30	Sequence 19, Appl1
11	11	33.8	12.2	2403	2	US-08-471-044-30	Sequence 30, Appl1
12	12	33.8	12.2	2403	2	US-08-463-483A-30	Sequence 30, Appl1
13	13	33.8	12.2	2403	2	US-08-471-046A-30	Sequence 30, Appl1
14	14	33.8	12.2	2403	2	US-08-470-566B-30	Sequence 30, Appl1
15	15	33.8	12.2	2403	2	US-08-838-2119B-7	Sequence 7, Appl1
16	16	33.8	12.2	2403	2	US-08-469-33A-30	Sequence 30, Appl1
17	17	33.8	12.2	2403	3	US-09-300-529-30	Sequence 30, Appl1
18	18	33.8	12.2	2403	3	US-09-233-336A-7	Sequence 7, Appl1
19	19	33.8	12.2	2403	3	US-09-233-752A-7	Sequence 7, Appl1
20	20	33.8	12.2	2403	4	US-09-402-036-7	Sequence 7, Appl1
21	21	33.6	12.1	688	5	PCT-US94-04361-26	Sequence 26, Appl1
C	22	33.4	12.0	777	4	US-08-959-116-218	Sequence 218, Appl1
C	23	33.4	12.0	2241	2	US-08-838-2119B-20	Sequence 20, Appl1
24	24	33.4	12.0	2241	3	US-09-233-536A-20	Sequence 20, Appl1
25	25	33.4	12.0	2241	3	US-09-233-752A-20	Sequence 20, Appl1
26	26	33.4	12.0	2241	4	US-09-402-036-20	Sequence 20, Appl1
27	27	33.4	12.0	2846	4	US-09-613-182-5	Sequence 5, Appl1

C	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
	3.3	3.4	4.1	3.3	3.3	3.3	3.1	3.1	3.1	3.1	3.1	3.1	3.1	3.1	3.1	3.1	3.1	3.1
	12.0	12.0	12.0	11.9	11.9	11.9	11.9	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4
	50341	50341	50341	349	2030	2030	2030	2576	2576	2576	2576	2576	2576	2576	2576	2576	2576	2576
	US-08-347-901C-1	US-08-347-901C-1	US-08-347-901C-1	US-08-275-925-13	US-08-530-950-11	US-08-888-429A-1	US-08-888-429A-1	US-08-471-033-35	US-08-471-033-35	US-08-471-033-35	US-08-471-033-35	US-08-471-033-35	US-08-471-033-35	US-08-471-033-35	US-08-471-033-35	US-08-471-033-35	US-08-471-033-35	US-08-471-033-35
	Sequence 1, Appl I	Sequence 1, Appl I	Sequence 1, Appl I	Sequence 13, Appl I	Sequence 1, Appl I	Sequence 1, Appl I	Sequence 1, Appl I	Sequence 35, Appl I	Sequence 35, Appl I	Sequence 35, Appl I	Sequence 35, Appl I	Sequence 35, Appl I	Sequence 35, Appl I	Sequence 35, Appl I	Sequence 26, Appl I	Sequence 17, Appl I	Sequence 26, Appl I	Sequence 17, Appl I

## ALIGNMENTS

```

1 RESULT
2 US-09-426-436 1/c
3 Sequence 1. Application US/09426436
4 Patent No. 6225066
5
6 GENERAL INFORMATION:
7
8 APPLICANT: William R. Jacobs, Jr.
9
10 APPLICANT: Barry R. Bloom
11
12 APPLICANT: Graham F. Hatfull
13
14 TITLE OF INVENTION: MYCOBACTERIAL SPECIES-SPECIFIC
15
16 TITLE OF INVENTION: REPORTER MYCOBACTERIOPHAGES
17
18 NUMBER OF SEQUENCES: 1
19
20 CORRESPONDENCE ADDRESS:
21
22 ADDRESSEE: Amster, Rothstein & Eidenstein
23
24 STREET: 90 Park Avenue
25
26 CITY: New York
27
28 STATE: New York
29
30 COUNTRY: U.S.A.
31
32 ZIP: 10016
33
34 COMPUTER READABLE FORM:
35
36 MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
37
38 COMPUTER: IBM PC Compatible
39
40 OPERATING SYSTEM: MS-DOS
41
42 SOFTWARE: Word Processor (ASCII)
43
44 CURRENT APPLICATION DATA:
45
46 APPLICATION NUMBER: US/09/426,436
47
48 FILING DATE:
49
50 CLASSIFICATION:
51
52 PRIOR APPLICATION DATA:
53
54 APPLICATION NUMBER: US/08/705,557
55
56 FILING DATE:
57
58 APPLICATION NUMBER: US/08/057,531
59
60 FILING DATE:
61
62 APPLICATION NUMBER: 07/833,431
63
64 FILING DATE: February 7, 1992
65
66 ATTORNEY/AGENT INFORMATION:
67
68 NAME: Pasquallini, Patricia A.
69
70 REGISTRATION NUMBER: 34,894
71
72 REFERENCE/DOCKET NUMBER: 96700/238
73
74 TELECOMMUNICATION INFORMATION:
75
76 TELEPHONE: (212) 697-5995
77
78 TELEFAX: (212) 286-0854 or 286-0082
79
80 TELETX: TWX 710-561-4766
81
82 INFORMATION FOR SEQ ID NO: 1:
83
84 SEQUENCE CHARACTERISTICS:
85
86 LENGTH: 52297
87
88 TYPE: nucleotide
89
90 STRANDEDNESS: single
91
92 TOPOLOGY: linear
93
94 MOLECULE TYPE:
95
96 DESCRIPTION: phage genome sequence
97

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 15, 2002, 09:47:03 ; Search time 2032.22 Seconds  
(without alignments)  
2862.671 Million cell updates/sec

Title: US-09-445-362b-6

Perfect score: 278  
Sequence: 1 gccacacagcancctcgagga.....catttnaactcccgagacc 278

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 segs, 10463268293 residues  
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: GenEmbl:\*  
2: gb\_ba:\*  
3: gb\_htg:\*  
4: gb\_in:\*  
5: gb\_om:\*  
6: gb\_ov:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_on:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	265.6	95.5	187532	9	AC006333	AC006333 Homo sapi
2	264	95.0	1936	6	AX002051	AX002051 Sequence
3	264	95.0	2080	6	AX002052	AX002052 Sequence
4	264	95.0	2268	6	AX002053	AX002053 Sequence
5	261	93.9	279	6	AX002056	AX002056 Sequence
6	216	77.7	168526	2	AC103173	AC103173 Rattus no
7	214.4	77.1	1458	10	AF237628	AF237628 Mus muscu
8	152	54.7	21542	9	AL513217	AL513217 Human DNA
9	152	54.7	171423	2	AC025664	AC025664 Homo sapi
10	150.4	54.1	3881	6	AX329789	AX329789 Sequence
11	150.4	54.1	3881	6	AX331533	AX331533 Sequence
12	150.4	54.1	3881	6	AX331932	AX331932 Sequence
13	150.4	54.1	3881	6	AX335915	AX335915 Sequence
14	150.4	54.1	3881	6	HSALTAN64	X54162 Human mRNA
15	150.4	54.1	228201	2	AC016814	AC016814 Mus muscu
16	137.6	49.5	1760	9	AK057852	AK057852 Homo sapi
17	137.6	49.5	182151	2	AC092060	AC092060 Homo sapi
18	131.2	47.2	518	11	G56351	G56351 SHGC-101709
19	120.8	43.5	93276	2	AC096239	AC096239 Rattus no
20	118.8	42.7	1572	10	S76831	S76831 Tmod-tropo
21	115.6	41.6	1837	10	AF237629	AF237629 Mus muscu
22	114	41.0	1550	10	AF177170	AF177170 Mus muscu
23	110.4	39.7	1190	10	RN059240	US9240 Rattus norv
24	109.2	39.3	1353	10	RN059241	US9241 Rattus norv
25	106	38.1	1067	5	AB052132	AB052132 Gallus ga
26	104.4	37.6	2493	5	CHKTMOD1A	L36678 Gallus dome
27	104	37.4	2491	9	AF177169	AF177169 Homo sapi
28	101.8	36.6	1780	9	BC002660	BC002660 Homo sapi
29	101.8	36.6	2865	9	HUMTRMOD	M77016 Human troph
30	94.4	34.0	1584	9	AF177171	AF177171 Homo sapi
31	94.4	34.0	2072	9	AF237631	AF237631 Homo sapi
32	94.4	34.0	2142	9	BC020542	BC020542 Homo sapi
33	89.6	32.2	1806	4	AF237632	AF237632 Sus scrof
34	88.4	31.8	1231	5	AF165215	AF165215 Gallus ga
35	88	31.7	1442	5	AF165216	AF165216 Dantio rer
36	86.8	31.2	1001	5	AB052717	AB052717 Gallus ga
37	84.8	30.5	3540	10	AF237630	AF237630 Mus muscu
38	83.2	29.9	1764	10	AF177172	AF177172 Mus muscu
39	73.4	26.4	1161	10	AF177174	AF177174 Homo sapi
40	70.2	25.3	1260	9	BC017810	BC017810 Homo sapi
41	70.2	25.3	1267	9	AF177173	AF177173 Homo sapi
42	68.6	24.7	872	10	AF28773857	AF287744 Mus muscu
43	68.6	24.7	1235	9	AF165217	AF165217 Homo sapi
44	68	24.5	3551	9	HSM802280	AL137543 Homo sapi
45	61.2	22.0	162863	2	AC096430	AC096430 Rattus no

## ALIGNMENTS

RESULT 1	AC006333	187532 bp	DNA	linear	PRI 30-SEP-2000,
LOCUS	AC006333	Homo sapiens BAC clone RP11-390E23	from 7,	complete sequence.	
DEFINITION	AC006333	Homo sapiens BAC clone RP11-390E23	from 7,	complete sequence.	
ACCESSION	AC006333	GI:5523811			
VERSION	AC006333.3	GI:5523811			
KEYWORDS	HTG.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
TITLE	1 (bases 1 to 187532)				
JOURNAL	Sulston,J.E. and Waterston,R.				
AUTHORS	Toward a complete human genome sequence				
REFERENCE	Genome Res. 8 (11), 1097-1108 (1998) .				
MEDLINE	99063792				
2 (bases 1 to 187532)					
REFERENCE	Rohlfing,T., Bauer,C. and Ketterman,M.				
AUTHORS	The sequence of Homo sapiens BAC clone RP11-390E23				
TITLE	3 (bases 1 to 187532)				
JOURNAL	Unpublished				
REFERENCE	Waterston,R.H.				
AUTHORS	Direct Submission				
TITLE					

JOURNAL Submitted (09-JAN-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 4 (bases 1 to 187532)

AUTHORS Waterston,R.H.

TITLE Direct Submission

JOURNAL Submitted (17-JUL-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 5 (bases 1 to 187532)

AUTHORS Waterston,R.

TITLE Direct Submission

JOURNAL Submitted (30-SEP-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

COMMENT On Jul 17, 1999 this sequence version replaced g1:4434617.

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>

Contact: [saplens@wustl.edu](mailto:saplens@wustl.edu)

----- Summary Statistics

Center project name: H\_NH0390E23

-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

\* This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

\* MAPING INFORMATION:  
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa,K., Moon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanesi,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://Dacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the right is a GSI-195F7, 200 bp overlap. Actual start is at base position 1 of RPI1-390E23; actual end is at 1716 of GSI-195F7.

The sequence fidelity is in question at base positions 75394-75555 due to a non-perfect GA run. The insilico digest fragments are 3570 for ecorv and 8922 for hindIII compared to 3573 for ecorv and 8997 hindIII in the real fragment digest. Assembly appears to be correct according to the restriction digests.

FEATURES

source

1..187532

Location/Qualifiers

1..187532

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="7"

/map="7"

repeat\_region /clone="RPI1-390E23"

repeat\_region /clone\_lib="RPCI-11"

repeat\_region 2..182

repeat\_region /rpt\_family="Retroviral"

repeat\_region 190..231

repeat\_region /rpt\_family="(CA)n"

repeat\_region 327..386

repeat\_region /rpt\_family="MALR"

repeat\_region 392..526

repeat\_region /rpt\_family="Alu"

repeat\_region 527..834

repeat\_region /rpt\_family="Alu"

repeat\_region 836..1227

repeat\_region /rpt\_family="MALR"

repeat\_region 3024..3337

repeat\_region /rpt\_family="Alu"

repeat\_region 3856..4008

repeat\_region /rpt\_family="L2"

repeat\_region 4823..5128

repeat\_region /rpt\_family="Alu"

repeat\_region 5696..5775

repeat\_region /rpt\_family="MIR"

repeat\_region 6130..6323

repeat\_region /note="match to EST AA984773 (NID:g3163298) am90n03.s1"

repeat\_region 6433..6524

repeat\_region /rpt\_family="L1"

repeat\_region 6591..6880

repeat\_region /rpt\_family="Alu"

repeat\_region 6925..7231

repeat\_region /rpt\_family="Alu"

repeat\_region 7258..7561

repeat\_region /rpt\_family="L1"

repeat\_region 8101..8622

repeat\_region /note="match to EST AA716350 (NID:g2728624) zh29d08.s1"

repeat\_region 9122..9332

repeat\_region /rpt\_family="MER1\_type"

repeat\_region 10506..10842

repeat\_region /rpt\_family="Other"

repeat\_region 10846..10881

repeat\_region /rpt\_family="(TA)n"

repeat\_region 10912..10940

repeat\_region /rpt\_family="(TTTAA)n"

repeat\_region 12699..12908

repeat\_region /rpt\_family="L1"

repeat\_region 13697..13718

repeat\_region /rpt\_family="(TTTAA)n"

repeat\_region 13719..14081

repeat\_region /rpt\_family="L1"

repeat\_region 15249..15401

repeat\_region /rpt\_family="MIR"

repeat\_region 15380..15568

repeat\_region /rpt\_family="MER1\_type?"

repeat\_region 17375..17504

repeat\_region /rpt\_family="L2"

repeat\_region 17612..17661

repeat\_region /rpt\_family="L2"

repeat\_region 17774..17796

repeat\_region /rpt\_family="AT-rich"

repeat\_region 17905..18016

repeat\_region /rpt\_family="Alu"

repeat\_region 18017..18048

repeat\_region /rpt\_family="(TAA)n"

repeat\_region 18650..19009

repeat\_region /rpt\_family="L2"

repeat\_region 19010..19133

repeat\_region /rpt\_family="Alu"

repeat\_region 19166..19254

repeat\_region /rpt\_family="MER1\_type"

repeat\_region 19383..19687

repeat\_region /rpt\_family="L1"

repeat\_region 20281..20428

repeat\_region /rpt\_family="L1"

repeat\_region 20429..20456

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: June 13, 2002, 09:45:35 ; Search time 30.86 Seconds  
(without alignments)  
3094.401 Million cell updates/sec

Title: US-09-445-362b-4  
Perfect score: 2839  
Sequence: 1 MSLPGRRRLSKYKESIDDE.....SIKOLKRVVPEALKWEHDL 552

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SPREMBL\_19:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_mhc:\*  
9: sp\_organelle:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1536	54.1	333	11 099PM7	099PM7 mus musculus
2	679	23.9	348	4 096LS4	096LS4 homo sapien
3	656.5	23.1	351	11 09UKK7	09UKK7 mus musculus
4	656.5	23.1	351	11 P70566	P70566 rattus norv
5	656.5	23.1	359	11 P70567	P70567 rattus norv
6	650.5	22.9	351	4 09NZR1	09NZR1 homo sapien
7	650.5	22.9	351	11 09JH9	09JH9 mus musculus
8	650.5	22.9	359	11 09ERR9	09ERR9 mus musculus
9	644.5	22.7	359	13 091006	091006 gallus galli
10	627.5	22.1	359	4 09BUT1	09BUT1 homo sapien
11	614	21.6	352	4 09NZR0	09NZR0 homo sapien
12	613	21.6	352	4 09NYL9	09NYL9 homo sapien
13	609	21.5	352	6 09NOY9	09NOY9 sus scrofa
14	606.5	21.4	345	13 09DEA6	09DEA6 gallus galli
15	590	20.8	345	11 09JH0	09JH0 mus musculus
16	561	19.8	345	4 09NZO9	09NZO9 homo sapien

17	560	19.7	345	11 09JH8	09JH8 mus musculus
18	555	19.5	345	4 09UKH2	09UKH2 homo sapien
19	546.5	19.2	343	13 09PU06	09PU06 brachydanio
20	541	19.1	347	13 09PU07	09PU07 gallus galli
21	517	18.2	333	13 09DEA4	09DEA4 gallus galli
22	382.5	13.5	392	5 001479	001479 caenorhabdi
23	374	13.2	367	5 046231	046231 drosophila
24	374	13.2	403	5 09VAS6	09VAS6 drosophila
25	369	13.0	401	5 095088	095088 caenorhabdi
26	248.5	8.8	269	4 09BV00	09BV00 homo sapien
27	242.5	8.5	1077	11 09JHCL	09JHCL mus musculus
28	240.5	8.5	737	4 09Y2L7	09Y2L7 homo sapien
29	240.5	8.5	1013	4 09NT81	09NT81 homo sapien
30	228.5	8.0	1112	4 096PY5	096PY5 homo sapien
31	228	8.0	168	11 09D386	09D386 mus musculus
32	226.5	8.0	602	2 09AKP3	09AKP3 rickettsia
33	226.5	8.0	1179	12 091L36	091L36 white spot
34	225.5	7.9	4833	11 09QYX6	09QYX6 mus musculus
35	225.5	7.9	5038	11 09QYX7	09QYX7 mus musculus
36	224.5	7.9	240	11 09CUR4	09CUR4 mus musculus
37	212	7.5	89	4 09NT43	09NT43 homo sapien
38	212	7.5	3394	4 09Y6V0	09Y6V0 homo sapien
39	211.5	7.4	360	4 09H985	09H985 homo sapien
40	211.5	7.4	1520	4 015087	015087 homo sapien
41	211.5	7.4	1781	4 09UKX0	09UKX0 homo sapien
42	211.5	7.4	1890	4 09UKW3	09UKW3 homo sapien
43	211.5	7.4	2073	4 09UKW2	09UKW2 homo sapien
44	211	7.4	542	12 011448	011448 anagrapa f
45	211	7.4	1110	13 091255	091255 petromyzon

## ALIGNMENTS

RESULT ID	1	PRELIMINARY:	PRT:	333 AA.
Q99PM7	Q99PM7			
AC	Q99PM7			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	CARDIAC LEIOMODIN (FRAGMENT).			
GN	LMOD2.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-21218919; PubMed-11318603;			
RA	Conley C.A., Fritz-Six K.L., Almenar-Queral A., Fowler V.M.;			
RT	"Leiomodins: larger members of the tropomodulin (tmcd) gene family.";			
RL	Genomics 73:127-139(2001).			
DR	EMBL; AF237628; AK00789.1; -.			
DR	MED; MGI:2135672; Lmod2.			
FT	NON_TER			
FT	NON_TER			
SO	SEQUENCE	333 AA;	37463 MW;	F082494410157A31 CRC64;

QY	214	TEVNIINNTENTTTOTLTFREFAALKDNTVTFSLANTHADDSAMAIAEMIKANEHTVY 273	54.1%;	Score 1536;	DB 11;	Length 333;
Db	4	TEVNIINNTENTTTOTLTFREFAALKDNTVTFSLANTHADDSAMAIAEMIKANEHTSV 63	90.1%;	Pred. No. 1.4e-75;		
QY	274	NVESNFTGKGLIALMALOHNTVLELRFHNRHNGSOVEEIKYKLENTTLRLG 333	13;	Mismatches 16;	Indels 4;	Gaps 2;
Db	64	NVESNFTGKGLIALMALOHNTVLELRFHNRHNGSOVEEIKYKLENTTLRLG 123				
QY	334	HFELPGMSMTSLTRMKOKROKLOEOKOCGCGGGLNLRKKVORGTSPSSPVSP 393				
Db		HFELPGMSMTSLTRMKOKROKLOEOKOCGCGGGLNLRKKVORGTSPSSPVSP 393				

Db 124 HFEELPGRMSMTSILTRNMDKQKQKMOEQKQEGHDGALRTKVMQGTGSSPYASP 183  
QY 394 RHSPSSPKLPKVVQTVRSPLSPVATLPPPPPPPPPPSSQRLPPPPPPPLPEKK 453  
Db 184 ROSPPSSPKVKKVHTGRSPSPVA--PPPPPPPLPP--HMLPPPPPPAPLPPEKK 239  
QY 454 LITRNAEVIKQOESQORALONGQKKKKKKYKOPNSTIKELKNSLRVQKKMEDSR 513  
Db 240 LITRNAEVIKQOESQORALONGQKKKKKKYKOPNSTIKELKNSLRVQKKMEDSR 299  
QY 514 PSTPORSAHENLMEALRGSSIKOLKREVEPALR 547  
Db 300 PSTPORSAHENLMEALRGSSIRQLRREVEPALR 333  
RESULT 2  
QY 09JUK7 PRELIMINARY: PRT: 348 AA.  
ID 096LS4  
AC 096LS4;  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE CDNA FLJ25123 FIS, CLONE CBR06154.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRN;  
RA Tashiro H., Yamazaki M., Matanabe K., Kumagai A., Itakura S.,  
RA Fukuzumi Y., Fujimori Y., Komiyama M., Suzuki Y., Hata H.,  
RA Nakagawa K., Mizuno S., Morinaga M., Kawamura M., Sugiyama T.,  
RA Irie R., Otsuki T., Sato H., Nishikawa T., Sugiyama A., Kawakami B.,  
RA Nagai K., Isogai T., Sugano S.;  
RT "NEO human cDNA sequencing project."  
RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AK051852; BAB71596.1; -  
SQ SEQUENCE 348 AA: 40194 MW: 6005E271D460B087 CRC64;

Query Match 23.9%; Score 679; DB 4; Length 348;  
Best Local Similarity 45.2%; Pred. No. 2,3e-29;  
Matches 141; Conservative 53; Mismatches 102; Indels 16; Gaps 2;

QY 72 YWEKESQKLEKRGEGCKVADEKESFEELFTESSEVSEYTTTEEBEESOEDEE 131  
Db 2 YWEKASRRRLKEERVPVTFVKSEELNNEIYANKRESKSSNIOETDEDEEDDDDD 61  
QY 133 EDSDEERTIETAKNGINGTVNDVSNDSN---KPIFKSQIENINLTNGSNGRTE-- 185  
Db 63 EEDDGESEETNREBECKAKQINCENCOQVTDKAFKEORDRPEAQOESKKISKLD 121  
QY 186 -----SPAIIHPCNPTVIDALDKISNDPDTTEVNLNENITQTTLRFAEA 235  
Db 122 PKLALDTSFLKVTSPSGNQTDLDGSLRVKNDPDKELNINNIENIPREMLIDFNA 181  
QY 236 LKDNVTVTFSLANTHADDSAAAMAIAEMLKANEHTNNVNSNFTTGCIATIMRALOH 295  
Db 192 MKKNHITFSLANGADEBNVAFALANMLRENRSITTLINISNFTTGCIYAIMRCLOFN 241  
QY 296 TVLTSLRFNORHINGSOVEMEIYKLEKNTTLRLGTHFELPGRMSMTSILTRNMDK 355  
Db 242 ETLTLRFRHNOHMLGHAHEIARLLKANNTLKMVGHFELPGRMVVNTLTRNDKQ 301  
QY 356 ROKRLQEQKQ 367  
Db 302 ROKRQEQKQ 313

RESULT 3  
QYJUK7 PRELIMINARY: PRT: 351 AA.  
ID 09JUK7

AC 09JUK7;  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE NEURAL TROPOMODULIN N-TMOD.  
GN TMOD2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Conley C.A., Almenar-Queralt A., Fowler V.M.,  
RT "Identifying novel tropomodulin isoforms."  
RL Mol. Biol. Cell 9:18A-18A(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Conley C.A., Fowler V.M.,  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF237629; AAF45297.1; -  
DR MGI: 1355335; Tmod2.  
SQ SEQUENCE 351 AA: 39510 MW: D01FDDEE185C828F CRC64;

Query Match 23.1%; Score 656.5; DB 11; Length 351;  
Best Local Similarity 39.9%; Pred. No. 3.7e-28;  
Matches 148; Conservative 72; Mismatches 116; Indels 35; Gaps 8;

QY 6 YRGRSKVESIDDELLASIAEELKELRELEDEPD-RNLVGLRQKSLTEKTPGT 64  
Db 5 FOKLEKTKNIDDELGLKISEELKOLENVLDLDESATLPAGFRQKQTKAATGPP 64  
QY 65 SREALMAYWENESQKLEKRGEGCKVADEKESFEELFT-ESNSEV-----SEYITE 119  
Db 65 DREHLMYLKE-----ALEQKDREDFVFTEGKGRVITPREKVEVR 108  
QY 120 EEESESEDEEE-----DSDEERTITANGINGTVNDVSNDSNPKPIFKSQIENINL 175  
Db 109 KEKVYTLDPLEELASASDTELYDLAAVLGVNHLNPNKFDDETTNGGGRKGPVRNV-- 166  
QY 176 TNGSNGRNT-ESPAIIHPCNPTVIDALDKISNDPDTTEVNLNENITQTTLRFAE 234  
Db 167 VKGSKAPVEDEPP-----NPTNVEASLQOMKANDPSLOEVNLNINIKPIPTLKEPAK 220  
QY 235 ALKDNVTVTFSLANTHADDSAAAMAIAEMLKANEHTNNVNSNFTTGCIATIMRALOH 294  
Db 221 SLENTIYVKKFSLAATRSNDPVALAEFAEMLVKNTKLSLVESNFTTGCIATLVALRE 280  
QY 295 TVLTSLRFNORHINGSOVEMEIYKLEKNTTLRLGTHFELPGRMSMTSILTRNMDK 354  
Db 281 NDTLTETKIDNQRQOJSTAVEKEIADWLERNRILKFGYQFTQKGPRTVAAATIKNNDL 340  
QY 355 QROKRLQEQK 365  
Db 341 VRKKRVGDRR 351

RESULT 4  
P70566  
ID P70566 PRELIMINARY: PRT: 351 AA.  
AC P70566;  
DT 01-FEB-1997 (Tremblrel. 02, Created)  
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE N-TROPOMODULIN.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxId=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRN;  
RX MEDLINE=97041109; PubMed=8886980;

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2002, 09:45:55 ; Search time 13.45 Seconds

(without alignments)  
1589,085 Million cell updates/sec

Title: US-09-445-362b-4

Perfect score: 2839

Sequence: 1 MSTFGYRGLSKYESIDEDE.....SIKOLKRVPEALRWEMHDL 552

Scoring table: BL0SUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	914	32.2	572	1 LMD1_HUMAN
2	650.5	22.9	359	1 TMOD_MOUSE
3	631.5	22.2	359	1 TMOD_HUMAN
4	209	7.4	2004	1 MOZ_HUMAN
5	199	7.0	1953	1 BN11_YEAST
6	193.5	6.8	1375	1 BN11_YEAST
7	189.5	6.7	1790	1 SEPA_EMENI
8	185.5	6.5	474	1 VP61_NPYOP
9	182	6.4	485	1 SSGP_VOICA
10	182	6.4	802	1 NAB3_YEAST
11	182	6.4	1362	1 BRD4_HUMAN
12	181	6.4	543	1 VP61_NPVAC
13	180.5	6.4	1248	1 DIAL_HUMAN
14	180.5	6.4	2517	1 NCR2_HUMAN
15	178.5	6.3	415	1 ACR0_HUMAN
16	178	6.3	384	1 VASP_CANPA
17	177.5	6.3	2472	1 NCR2_MOUSE
18	176.5	6.2	431	1 ACR0_RABIT
19	174.5	6.1	421	1 ACR0_HUMAN
20	173.5	6.1	1101	1 DIAL_HUMAN
21	172	6.1	1206	1 G3PT_MOUSE
22	172	6.1	1206	1 FMI4_MOUSE
23	171	6.0	380	1 VASP_HUMAN
24	171	6.0	771	1 YXCO_YEAST
25	171	6.0	1230	1 ST20_CANAL
26	170.5	6.0	451	1 MYBH_DICDI
27	165	5.8	141	1 VPRO_OMERU
28	165	5.8	1091	1 DIAL_DROME
29	164	5.8	1468	1 FMI1_MOUSE
30	164	5.8	1781	1 AKAC_HUMAN
31	163.5	5.8	2774	1 NAPA_RAT
32	163	5.7	501	1 WASL_RAT
33	163	5.7	1213	1 FMN_CHICK

34	163	5.7	1574	1 SYJ1_RAT	Q62910 rattus norv
35	163	5.7	1875	1 MLPI_YEAST	Q02455 saccharomyc
36	162.5	5.7	505	1 WASL_HUMAN	Q00401 homo sapien
37	162	5.7	505	1 WASL_BOVIN	Q95107 bos taurus
38	162	5.7	3924	1 ANK2_HUMAN	Q01484 homo sapien
39	161.5	5.7	3726	1 ABFI_MOUSE	Q61329 mus musculu
40	161	5.7	681	1 MP10_HUMAN	Q00566 homo sapien
41	161	5.7	1018	1 FNBA_STPAU	P14738 staphylococ
42	160.5	5.7	487	1 EBN2_EBV	P12978 Epstein-Bar
43	159	5.6	304	1 CEC1_CAEEL	P34618 caenorhabdi
44	159	5.6	2442	1 CBR_HUMAN	Q92793 homo sapien
45	158.5	5.6	379	1 YP85_CAEEL	Q09442 caenorhabdi

## ALIGNMENTS

RESULT 1  
LMD1\_HUMAN STANDARD; PRT; 572 AA.  
AC P29536;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE leiomodlin 1 (leiomodlin, muscle form) (64 kDa autoantigen D1) (64 kDa autoantigen ID) (64 kDa autoantigen ID3) (Thyroid-associated opthalmopathy autoantigen) (Smooth muscle leiomodlin) (SM-Lmod).  
GN LMOD1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Thyroid;  
RX MEDLINE=91225220; PubMed=2026759;  
RA Dong Q., Ludgate M., Vassart G.;  
RT "Cloning and sequencing of a novel 64-kDa autoantigen recognized by patients with autoimmune thyroid disease."  
RL J. Clin. Endocrinol. Metab. 72:1375-1381(1991).  
RN [2]  
RP CHARACTERIZATION.  
RX MEDLINE=99451105; PubMed=10520227;  
RA Conley C.A., Fowler V.M.;  
RT "Localization of the human 64kd autoantigen D1 to myofibrils in a subset of extraocular muscle fibers."  
RL Curr. Eye Res. 19:313-322(1999).  
RN [3]  
RP CHARACTERIZATION.  
RX MEDLINE=21248187; PubMed=11350761;  
RA Conley C.A.;  
RT "leiomodlin and tropomodulin in smooth muscle."  
RL Am. J. Physiol. 280:C1645-C1656(2001).  
RN [4]  
RP CHARACTERIZATION.  
RX MEDLINE=21218919; PubMed=11318603;  
RA Conley C.A., Fritz-Six K.L., Almenar-Queralt A., Fowler V.M.;  
RT "leiomodins: larger members of the tropomodulin (Tmod) gene family."  
RL Genomics 73:127-139(2001).  
CC -1- SUBCELLULAR LOCATION: CYTOSKELETON.  
CC -1- TISSUE SPECIFICITY: SMOOTH MUSCLE (HEART, SKELETAL MUSCLE, COLON AND SMALL INTESTINE), A SUBSET OF STRIATED MUSCLE FIBERS, AND AT LOW LEVEL IN THYROID.  
CC -1- DISEASE: RECOGNIZED BY PATIENTS WITH AUTOIMMUNE THYROID DISEASE.  
CC -1- SIMILARITY: BELONGS TO THE TROPOMODULIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to [license@sdb.ch](mailto:license@sdb.ch)).

CC	EMBL: X54162; CAA38101.1; -. PIR: S26815. S26815. MIM: 602715. InterPro: IPR003124; WH2. Pfam: PF02005; WH2; 1. SMART: SM00246; WH2; 1. Kw Antigen: Repeat; Cytoskeleton. FT DOMAIN 137 265 REPEAT 137 152 REPEAT 153 168 REPEAT 169 184 REPEAT 185 200 REPEAT 201 216 REPEAT 217 232 REPEAT 233 248 REPEAT 249 265 FT REPEAT 480 499 FT DOMAIN 572 AA; 63737 MM; SEQUENCE	5 8 1 2 3 4 5 6 7 8 5 X 4 AA DDB42P8E0523DE94 CnC64;	APPROXIMATE TANDEM REPEATS.
DR	EMBL: X54162; CAA38101.1; -. PIR: S26815. S26815. MIM: 602715. InterPro: IPR003124; WH2. Pfam: PF02005; WH2; 1. SMART: SM00246; WH2; 1. Kw Antigen: Repeat; Cytoskeleton. FT DOMAIN 137 265 REPEAT 137 152 REPEAT 153 168 REPEAT 169 184 REPEAT 185 200 REPEAT 201 216 REPEAT 217 232 REPEAT 233 248 REPEAT 249 265 FT REPEAT 480 499 FT DOMAIN 572 AA; 63737 MM; SEQUENCE	5 8 1 2 3 4 5 6 7 8 5 X 4 AA DDB42P8E0523DE94 CnC64;	APPROXIMATE TANDEM REPEATS.

Query Match	32.2%;	Score 914;	DB 1;	Length 572;
Best Local Similarity	34.2%;	Pred. No. 8.3e-35;		
Matches 220;	Conservative 91;	Mismatches 135;	Indels 198;	Gaps 11

Qy	85	-----	RLGCGK-----	91	
Db	61	SKOYETKTDAGKNGERDASKKALGPRRNSDLGEPKRGCLKKSFSRDRDAGGKSGEK	120		
Qy	92	-----	VAEDKKEES-----	EBELLFETSENSV-----	113
Db	121	PKEEKIIIRGIDKGRVAAVDKKEAGKMGDGRGEERAVATKEEKKGGDRNTGLSRDKKR	180		
Qy	114	-----	EEVYTEEEBESGOEEBEEBDSDEERTIETAG-----	INGVANTDSVN-----	157
Db	181	EEKKEVAKKEDDEKRVGGERNDTRKREGSKMRAAGGNTDMKEDBEKKRTGNTDKDD	240		
Qy	158	-----	SDNSKPKIIFKSOIENINLTJNGSNGRNTESPAAIHPCGNPTVIDAL	203	
Db	241	EKKVKNPEPLHEKAKDSDSKTKPEKO-----	TPSGPRTKPEEGPAKVEEBAEAPFIDEPL	294	
Qy	204	DKISNPDPTENVLNENNIENTTOTUTTPRALBKJNTVYAKFSLANHHADSAAMADEM	263		
Db	295	ERVKNNDPEMTBVVNNNSDCIETNEILVYRTTELLENYVYKLFALANTRADHDHAFALAIM	354		
Qy	264	LKANENITVNVVNSNFTTGKGLIATMRALQHTVYTELRFPHNOHNGMSOYEMETVYKLG	323		
Db	355	LKANAKTTITJNDSDNITGKGLIAFRLALQNTLTTELRFNORHIGOGKEMETIAKLK	414		
Qy	324	ENTTLRLYGHFELPPPKMSMTSILTRNMDKOROKRLEOQKOEGYOGGPRFKYVQNG	383		
Db	415	ENMSILKLGHYHFLAPARPTVTNLTLSRNMDKOROKRLEOQAOEANGKEKKDLLEVPAKG	474		
Qy	384	TPSSSPVPSVRHRSWSSPKLKKVQTVNSRPLSPVATLPPPPPPPPPPSSQRLPPPP	443		
Db	475	AVAKG-----	SKRPSPOSPK-----	ESPKNSPKGGAPAPAP	507
Qy	444	PPPPPLEPEKKLIRINIAEYIKOESQORALONGOKKKKKKKYKKOPNLSILKEINISRSV	503		
Db	508	PPPPPL-----	-----	APPLIMENLKNLSISPA	529
Qy	504	QEKKMESSSPSTPQSAHENTMEAIRGSSIKOLKRVVPEALR	547		
Db	530	TORKMGDKVLPRA-QEKNNSDOLLAARSSNLIKOLKVVVPPKLO	572		

RESULT	2	
TMOD_MOUSE		
ID	TMOD_MOUSE	STANDARD;
		PRT;
		359 AA

DT P49813.1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Tropomodulin.  
 DE TMOD OR TMOD1.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI-TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=95154574; PubMed=7851652;  
 RX Ito M., Swanson B., Sussman M.A., Kedes L., Lyons G.;  
 RT "Cloning of tropomodulin cDNA and localization of gene transcripts  
 RT during mouse embryogenesis";  
 RL Dev. Biol. 167:317-328(1995).  
 CC -I- FUNCTION: TROPOMYOSIN-REGULATING PROTEIN. IT MAY MODULATE THE  
 CC ASSOCIATION OF TROPOMYOSIN WITH THE SPECTRIN-ACTIN COMPLEX IN  
 CC THE ERYTHROCYTE MEMBRANE SKELETON, AND THUS THE VISCOCLASTIC  
 CC PROPERTIES OF ERYTHROCYTES. IT BINDS TO THE END OF ERYTHROCYTE  
 CC TROPOMYOSIN AND BLOCKS HEAD-TO-TAIL ASSOCIATION OF TROPOMYOSIN  
 CC ALONG ACTIN FILAMENTS (BY SIMILARITY).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>).  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: S76831; AAB3388.1; -;  
 DT MGD: MGI:98775; Tmod1  
 DT DOKAIN 39 138 TROPOMYOSIN-BINDING (POTENTIAL).  
 SE SEQUENCE 359 AA; 40484 MW; 5EE340B83046823D CRC64;

Query Match	22.9%	Score 650.5	DB 1	Length 359
Best Local Similarity	40.1%	Pred. No. 3.7e-23		
Matches 150; Conservative	66	Mismatches 111	Indels 47	Gaps 8

```

QY 6 YRRSKTESIEDDELLASLSAEELKELEREDIEDIPEORN -LPVGLROCKSLTEKPTGPF 64
Db 3 YRRELEKRDIDEDDILATLEEBELRTLENELEDELDPPALLPAELRODOOTTKAPGPF 62
QY 65 SREALMAWMEKESQOLLEKERL ---GE -CGKVADEKESEBELIFTESNSVSEBYTE 119
Db 63 KREELDLHLEKQAKFKROREDLVPYTGKRGKVMWPKQKPMOPVL -----ESVILE 113
QY 120 EEEESQEEEBEDSE-----EBRTLEAKINGVAVDVSNSDKSKIRKS 168
Db 114 PELEBALNASPAELCDIAALIGMHTLMSNOOYQALSSSVNNEGINS -VIKPTQYKP 172
QY 169 QIENINLNGSNGRTESPAIIHPCGNPTFEDADLKISNDPOTTEVNLNINEITQT 228
Db 173 VPD-----EED-----NSTDVEETLERIKNNPDELEEVNLNNIRNIPPT 212
QY 229 LTRFAEALKDNTVYKTFSLANTHADDSAMAIAEMIKANEHTIVNVVNSNFITGKILAI 288
Db 213 LKAYAEALKENSUYKFSIVGTRSDPAPFALAEWLKYNKYLKATLVESNFIISAGILRL 272
QY 289 MRALDHTVLELRFPHNORHIMGSOVEWEIYKLLKENTTLRLGLYHFEELPGFRMSSTIL 348
Db 273 VEALPHNTSIVLELKIDNOSOPLGKNVKEVMEIYMLKENTTLLEKGYHFPQOQGRLRASHAM 332
QY 349 TRNMCKOROKRLOE 362
Db - 333 MSNNDLVKRRRLAD 346

```

RESULT	3
TMOD_HUMAN	

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2002, 09:43:59 ; Search time 21.02 Seconds  
(without alignments)  
2523.374 Million cell updates/sec

Title: US-09-445-362b-4

Sequence: 1 MSTFGYRRRLSKYESTIDEDE.....SIKQLKRVVPALRWMDL 552

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 71:\*

1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	914	32.2	572	S18732	autoantigen, 64k -
2	650.5	22.9	359	T53091	tropomodulin - mou
3	644.5	22.7	359	A55463	tropomodulin - skel
4	631.5	22.2	359	A42336	tropomodulin - hum
5	310.5	10.9	324	T25521	hypothetical prote
6	240.5	8.5	1013	T46422	hypothetical prote
7	235	8.3	687	T29220	hypothetical prote
8	212	7.5	89	T46384	hypothetical prote
9	211	7.4	1110	T51116	NF-180 - sea lampy
10	203.5	7.2	1621	T15264	hypothetical prote
11	203	7.2	1461	T41643	probable involveme
12	202.5	7.1	520	E97813	WASP, N-WASP, MENA
13	199	7.0	1953	S63244	BNI1 protein - yea
14	194	6.8	1657	T19536	hypothetical prote
15	193.5	6.8	1375	S48375	hypothetical prote
16	192.5	6.8	969	T15446	hypothetical prote
17	191	6.7	742	F84643	hypothetical prote
18	190.5	6.7	1100	T30967	transcription acti
19	189.5	6.7	980	S54986	regulatory protein
20	185.5	6.5	474	T10271	capacid-associated
21	182.5	6.4	1110	T19673	hypothetical prote
22	182	6.4	485	A33647	sulfated surface g
23	182	6.4	802	S48529	NAB3 protein - yea
24	181	6.4	1634	T26517	hypothetical prote
25	181	6.4	2526	T20531	hypothetical prote
26	181	6.4	2722	T20532	hypothetical prote
27	181	6.4	2728	E88320	protein F07A1.6
28	180.5	6.3	2562	T14266	kin protein - chic
29	180	6.3	543	S25128	61k protein - Auto

30	180	6.3	555	2	T30349	structural protein
31	179.5	6.3	473	2	B85187	glycoprotein homol
32	178.5	6.3	415	1	A34170	acrosin (EC 3.4.21
33	178.5	6.3	505	2	B64560	poly E-rich protei
34	178	6.3	384	2	S51796	vasodilator stimu
35	176.5	6.2	431	2	S47538	acrosin (EC 3.4.21
36	176.5	6.2	1388	2	T00063	hypothetical prote
37	175.5	6.2	681	2	F85062	hypothetical prote
38	175	6.2	599	2	T10798	phenophorin-S - Vo
39	174.5	6.1	421	1	S11674	acrosin (EC 3.4.21
40	174	6.1	1392	2	T51947	probable transcrip
41	173	6.1	907	2	E96636	hypothetical prote
42	172	6.1	440	2	I49681	glyceroldehyde-3-p
43	172	6.1	1206	2	S24407	formin isoform IV
44	172	6.1	5170	2	T15348	hypothetical prote
45	171.5	6.0	1230	2	T18256	probable serine/th

## ALIGNMENTS

RESULT	1	
S18732	autoantigen, 64k - human	
C:Species: Homo sapiens (man)		
C:Date: 31-Dec-1993	#sequence_revision 02-Aug-1994	#text_change 17-Mar-1999
C:Accession: S18732; S26815		
R:Dong, Q., Ludgate, M., Vassart, G.		
J. Clin. Endocrinol. Metab. 72, 1375-1381, 1991		
A:Title: Cloning and sequencing of a novel 64-kDa autoantigen recognized by patients		
A:Reference number: S18732; M0ID:91225220		
A:Accession: S18732		
A:Status: nucleic acid sequence not shown		
A:Molecule type: mRNA		
A:Residues: 1-572 <DON1>		
A:Cross-references: EMBL:X54162		
R:Dong, Q.H.		
submitted to the EMBL Data Library, July 1990		
A:Reference number: S26815		
A:Accession: S26815		
A:Molecule type: mRNA		
A:Residues: 1-112, 'A', 114-572 <DON2>		
A:Cross-references: EMBL:X54162; NID:928968; PID:928969		
C:Keywords: glycoprotein		
F:25,310,358,386,416/Binding site: carbohydrate (Asn) (covalent) #status predicted		
Query Match	32.2%	Score 914; DB 2; Length 572;
Best Local Similarity	34.2%	Pred. No. 3.2e-35;
Matches 220; Conservative 91; Mismatches 135; Indels 198; Gaps 11;		
QY 30 LKELELEDIEPPDNLVGLRQKSLTEKPTGTFSEALMAWYKESOKLLEKE----- 84		
DB 1 MEELKELDVDPDGSVVGRLRQNRQTEKOSTGYNNRAMLNPEKEKTKLMQEMSMDE 60		
QY 85 -----RLGECGK----- 91		
DB 61 SKQVETKIDAKNGQERGDASKAKALGPRRNSDLGKEPAKGLKRSFSDRDESGKSGEK 120		
QY 92 -----VAEDKEES-----EEELFTSENSEVS----- 113		
DB 121 PKEEKIRIGIDKGRVAADVKKKAGKDGREGRAVAATKKEEKKGGDRNTLSRDKDKR 180		
QY 114 -----EAVTTEEESESESESESESEDEERTTETAK-----INGTVNYSVN- 157		
DB 181 EEMKEVAKKEDDEKVGKGRNDRTRKEGKKRAGGNTDMKKEDEKVGKRGNDTKDD 240		
QY 158 -----SDNSKPKTFKSQIENINLTNSNGRNTESPAIHPCGNPIYIEAL 203		
DB 241 EKVKRNEPLHEKARDSDSKTPERQ-----TPSGPTKPSGPAKVEEAPSLFDEPL 294		
QY 204 DKIKSNDPPTVEVNNINENITQTLTRPAEALKNQTVVKKFSLANTADSAAMATAEW 263		





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: June 13, 2002, 09:43:39 ; Search time 14.61 Seconds  
(Without alignments)  
922.856 Million cell updates/sec

Title: US-09-445-362B-4  
Perfect score: 2839  
Sequence: 1 MSTRGYRGLSKYESIDEDF.....SIKQKRYEPEALRWEHDL 552

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: Issued\_Patents\_AA:\*  
2: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backkille1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	197	6.9	1105	4	US-08-999-774A-2
2	182.5	6.4	1315	3	US-08-899-595-3
3	180.5	6.4	1248	4	US-09-080-897-2
4	180.5	6.4	1248	4	US-09-323-735-2
5	171	6.0	380	2	US-09-026-587-4
6	171	6.0	380	2	US-09-227-420-4
7	164	5.8	1780	1	US-08-769-309A-5
8	164	5.8	1780	3	US-08-994-570-5
9	158.5	5.6	418	2	US-09-026-587-1
10	158.5	5.6	418	2	US-09-227-420-1
11	158	5.6	2414	1	US-08-227-536-2
12	158	5.6	2414	5	PCT-US95-04682-2
13	156.5	5.5	513	4	US-09-041-886-28
14	156.5	5.5	530	4	US-09-041-886-29
15	156.5	5.5	552	4	US-09-041-886-30
16	156.5	5.5	589	4	US-09-041-886-31
17	156.5	5.5	3144	1	US-08-246-982A-6
18	156.5	5.5	3144	1	US-08-453-265-6
19	156.5	5.5	3144	2	US-08-457-273B-42
20	156.5	5.5	3144	4	US-08-556-419-21
21	156.5	5.5	3144	4	US-09-041-886-15
22	155.5	5.5	1255	2	US-09-080-897-4
23	155.5	5.5	1255	3	US-08-899-595-1
24	155.5	5.5	1255	4	US-09-323-735-4
25	152.5	5.4	393	2	US-09-026-587-3
26	152.5	5.4	393	2	US-09-227-420-3
27	151	5.3	3119	1	US-08-246-982A-16

28	151	5.3	3119	1	US-08-453-265-16	Sequence 16, Appl
29	149.5	5.3	905	2	US-08-574-959A-9	Sequence 9, Appl
30	149.5	5.3	905	4	US-09-357-014-9	Sequence 9, Appl
31	149.5	5.3	1135	2	US-08-574-959A-7	Sequence 7, Appl
32	149.5	5.3	1135	4	US-09-357-014-7	Sequence 7, Appl
33	149	5.2	323	4	US-09-029-213B-25	Sequence 25, Appl
34	149	5.2	1805	1	US-07-853-913-2	Sequence 2, Appl
35	149	5.2	1958	1	US-07-945-283-2	Sequence 2, Appl
36	149	5.2	2441	1	US-08-194-468-2	Sequence 2, Appl
37	149	5.2	2441	3	US-08-961-739-2	Sequence 2, Appl
38	147	5.2	281	2	US-08-810-453-2	Sequence 2, Appl
39	147	5.2	281	3	US-08-815-190A-2	Sequence 25, Appl
40	147	5.2	281	4	US-09-290-640-25	Sequence 25, Appl
41	147	5.2	281	4	US-09-479-524-3	Sequence 3, Appl
42	147	5.2	281	4	US-08-339-214-8	Sequence 8, Appl
43	147	5.2	281	4	US-08-339-214-30	Sequence 30, Appl
44	147	5.2	281	5	PCT-US95-00362-2	Sequence 2, Appl
45	147	5.2	976	4	US-09-104-324B-4	Sequence 4, Appl

## ALIGNMENTS

RESULT 1  
US-08-999-774A-2  
; Sequence 2, Application US/08999774A  
; Patent No. 6274312  
; GENERAL INFORMATION:  
; APPLICANT: Gish, Kurt C.  
; APPLICANT: Seghezzi, Wolfgang  
; APPLICANT: Shanahan, Frances  
; APPLICANT: Lees, Emma M.  
; APPLICANT: Mcclanahan, Terrill K.  
; TITLE OF INVENTION: Intracellular Regulatory Molecules;  
; TITLE OF INVENTION: Related Reagents  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DMAX Research Institute  
; STREET: 901 California Avenue  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/999,774A  
; FILING DATE: 10-DEC-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/032,818  
; FILING DATE: 11-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ching, Edwin P.  
; REGISTRATION NUMBER: 34,090  
; REFERENCE/DOCKET NUMBER: DX0646  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650)852-9196  
; TELEFAX: (650)496-1200  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1105 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-999-774A-2

Query Match 6.9%; Score 197; DB 4; Length 1105;  
Best Local Similarity 20.0%; Pred. No. 1e-05;

[illegible]

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2002, 09:41:49 ; Search time 34.22 Seconds  
(without alignments)  
1791.722 Million cell updates/sec

Title: US-09-445-362B-4

Perfect score: 2839  
Sequence: 1 MSTFGYRGRLSKYSIDEDL.....SINKLRVEVPEALRWEHDL 552

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A.Geneseq\_032802:\*

1: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:\*  
2: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:\*  
3: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:\*  
4: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:\*  
5: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:\*  
6: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:\*  
7: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:\*  
8: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:\*  
9: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:\*  
10: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:\*  
11: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:\*  
12: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:\*  
13: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:\*  
14: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:\*  
15: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:\*  
16: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:\*  
17: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:\*  
18: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:\*  
19: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:\*  
20: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT:\*  
21: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:\*  
22: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2839	100.0	552	20	AAW90172 Human heart muscle
2	2839	100.0	552	22	AAAB6278 Human DCMAG-1 prot
3	2063.5	72.7	531	22	AAAB19573 Human diagnostic a
4	1447	51.0	302	21	AAAB2655 Human ORFX ORF2419
5	964	34.0	600	22	AAAB8842 Human polypeptide
6	963	33.9	610	22	AAAB0628 Human polypeptide
7	779.5	27.5	358	21	AAAY5848 Autoantigen diagno
8	709.5	25.0	330	21	AAAY5846 Autoantigen diagno
9	705	24.8	168	21	AAAB3003 Human ORFX ORF2767
10	650.5	22.9	351	22	AAAB79201 Human protein SFO
11	613	21.6	352	22	AAAB95533 Human polypeptide

12	613	21.6	352	22	AAAB98824 Human tropomodulin
13	613	21.6	358	22	AAAB41319 Human polypeptide
14	561	19.8	368	21	AAAB43100 Human ORFX ORF2864
15	553.5	19.5	391	22	AAAB03836 Novel human secret
16	480	16.9	104	22	AAAB21876 Human cardiovascular
17	416	14.7	223	22	AAAB01815 Human heart muscle
18	378	13.3	92	20	AAAB90173 Human protein sequ
19	376	13.2	174	22	AAAB25794 Human protein sequ
20	374	13.2	367	22	ABAB6832 Drosophila melanog
21	374	13.2	403	22	ABAB5818 Drosophila melanog
22	360	12.7	273	22	AAAB11834 Human polypeptide
23	348	12.3	141	22	AAAB03835 Novel human secret
24	240.5	8.5	1017	22	AAAB40352 Human polypeptide
25	226	8.0	1174	22	AAAB50339 Shrimp white spot
26	211.5	7.4	560	22	AAAB95239 Human protein sequ
27	198.5	7.0	1162	22	ABAB62516 Drosophila melanog
28	197	6.9	1105	22	ABAB9325 Drosophila melanog
29	195.5	6.9	527	22	ABAB58220 Human intracellular
30	194.5	6.9	983	20	AAAY09513 Mouse JMY protein
31	194.5	6.9	3111	22	AAAB60327 Drosophila melanog
32	193.5	6.8	757	22	AAAB50172 Shrimp white spot
33	193	6.8	1700	22	ABAB64608 Drosophila melanog
34	187	6.6	635	19	AAAB50334 Green fluorescent
35	187	6.6	635	21	AAAY0781 EGFP-VASP fusion p
36	186	6.6	3536	22	ABAB5480 Drosophila melanog
37	184	6.5	748	22	AAAB58340 Human brain expres
38	184	6.5	3263	22	ABAB67210 Drosophila melanog
39	183.5	6.5	2819	22	AAAB35408 Human 07C327 gene
40	182.5	6.4	1315	20	AAAB76734 Human mdia Rho tar
41	182	6.4	100	21	AAAG00851 Human secreted pro
42	181	6.4	2518	21	AAAB40574 Human ORFX ORF338
43	180.5	6.4	473	21	AAAG24963 Arabidopsis thalia
44	180.5	6.4	1248	20	AAAY13464 Human diaphanous p
45	180.5	6.4	2724	22	ABG20119 Novel human diagno

ALIGNMENTS

RESULT 1	
ID	AAW90172 standard: Protein; 552 AA.
XX	
AC	AAW90172;
XX	
DT	16-MAR-1999 (first entry)
XX	
DE	Human heart muscle specific protein.
XX	
KW	Heart muscle; cardiac insufficiency; detection; interactor; modulator;
KW	heart disease; diagnostic; therapy.
XX	
OS	Homo sapiens.
XX	
PN	WO9656907-A1.
XX	
PD	17-DEC-1998.
XX	
PF	15-JUN-1998; 98WO-EP03584.
XX	
PR	13-JUN-1997; 97DE-1025186.
XX	
PA	(MED1-) MEDIGENE AG.
XX	
PI	Domdey H, Henkel T, Hofmann ME;
XX	
DR	WPI: 1999-060330/05.
XX	
DR	N-PDB; AAAY74263, AAAY74264, AAAY74365.
XX	
PT	New nucleic acid selectively expressed in insufficient cardiac
XX	muscle - for treatment and diagnosis of heart disease
PS	Claim 1; Fig 4; 58pp; German.

XX This sequence represents a fragment of a human heart muscle specific  
CC protein. This sequence is used for treating heart disease, especially  
CC cardiac insufficiency, and for detecting interactors (e.g. in a  
CC two-hybrid assay, identifying polypeptides that interact with the  
CC sequence, potential therapeutic agents, or compounds that modulate  
CC such interactions). They can also be used diagnostically, e.g. in  
CC standard amplification, Northern blotting or immunoassays. This sequence  
CC has been identified in human cardiac tissue cDNA libraries and is more  
CC strongly expressed in diseased tissue than in normal tissues, suggesting  
CC that it is causally related to cardiac insufficiency. The new materials  
CC provide specific and accurate diagnosis and therapy of heart disease.  
XX  
SQ Sequence 552 AA:

Query Match 100.0%; Score 2839; DB 20; Length 552;  
Best Local Similarity 100.0%; Pred. No. 1,9e-169;  
Matches 552; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MSTFGYRRLSKYESIDEDELLASAELEKELELEDIEDPDRNLPGVLRKSLTEKTP 60
DB 1 mstfgyrrlglskyesidedellasaeelekelelediedpdrnlpgvlrqsalktekp 60
QY 61 TGTFSREALMAYWEKESQKLEKERLGECKVAEDKESEEEELFTESNSEVESEVYTEE 120
DB 61 tgftrsrealmaywekesqkllkerlgecgkvaedkeeseeeellftesenveesevytee 120
QY 121 EEEESQEEEDSDSEERTETETAKGNGVYVNSDNDKPKIFKSOIENINLTNGSN 180
DB 121 eeeesqeeeedseedeertletakgngvtyvnsdndskpkifksqienlnltngsn 180
QY 181 GRNTESPAIHPCGNPTVIEDALDKIKSNDPDTTEVNININENTITQTITRFAALKDNT 240
DB 181 grntespaahpcgnptviedaldkiksndpdttevnininentitqtitrfaealkdnt 240
QY 241 VVKFTSLANTADDSAAAMIAEMLKANEHTTNVNESNFTGKGLAIRMALQHTVYTE 300
DB 241 vvkftslantaddsaamaiaemlkanehttnvnesnftgkgilaalmalqhtvlyte 300
QY 301 LRFHNRHIMSGVEMETVTKLKENTTLRLGYHFEELPGPRMSMTSLTRMDQOROKRL 360
DB 301 lrfhngrhimgsvemeltvklkenttllrlgyhfeelpgprmsmtsltrmdqyrqkrl 360
QY 361 QEOKQOEGYDGGPNLRTKVMQGRTPSSSPYVSPRHSFWSFKLPKKQVTVSRPLSPVAT 420
DB 361 qeqkqgegydggpnlrktvwqgrtpssspyyvsprrhsfwsfplpkkqvtrsrplspvat 420
QY 421 LPPPPPPPPPPSSQRLPPPPPPPPPLPEKKLITRNIAEVITKQESQALQNGQKKK 480
DB 421 lppppppppppssqrlppppppppplpekklltrniaevitkqesaqalngqgkkk 480
QY 481 KGRKVKKOPNSILKEIKNSLSRVOEKMEKEDSSRSTPORSAHENLMEAIRGSSIKOLKRV 540
DB 481 kgkvvkkgpnslkellknsllsrvoekmedssrstrpqrshenlmeairgssikqlkry 540
QY 541 EVPALRWEHDL 552
DB 541 evpealrwehdl 552
```

RESULT 2  
ID AAB86278 standard; protein: 552 AA.  
XX AAB86278:

XX 06-SEP-2001 (first entry)  
XX Human DCMAG-1 protein.  
XX DCMAG-1: human; dilative cardiomyopathy; cardiac muscle cell;  
XX cardioactive agent; heart disease; treatment.  
KM

XX OS Homo sapiens.  
XX WO200146388-A2.  
XX 28-JUN-2001.  
XX 21-DEC-2000; 2000WO-EP13101.  
XX 22-DEC-1999; 99DE-1062154.  
XX (MED1-) MEDIGENE AG.  
XX Roenicke V, Nave B, Henkel T;  
XX WPI; 2001-418054/44.  
XX New pathologically altered cardiomyocyte, useful for identifying and  
XX detecting cardioactive agents, produced by stimulating healthy cells  
XX with hormones and cytokines -  
XX Claim 15; Page 38-40; 40pp; German.

This invention describes a novel pathologically altered cardiac muscle  
cell (A), isolated from healthy cardiac tissue and/or cells by  
stimulation of isolated cells with suitable hormones (or analogs) and/or  
cytokines. (A) are used to identify or detect cardioactive agents,  
potentially useful for prevention or treatment of heart disease, also to  
study molecular changes that cause heart disease. This sequence  
represents the human DCMAG-1 protein, associated with dilative  
cardiomyopathy which is used to illustrate the method of the invention.

SQ Sequence 552 AA:

Query Match 100.0%; Score 2839; DB 22; Length 552;  
Best Local Similarity 100.0%; Pred. No. 1,9e-169;  
Matches 552; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MSTFGYRRLSKYESIDEDELLASAELEKELELEDIEDPDRNLPGVLRKSLTEKTP 60
DB 1 mstfgyrrlglskyesidedellasaeelekelelediedpdrnlpgvlrqsalktekp 60
QY 61 TGTFSREALMAYWEKESQKLEKERLGECKVAEDKESEEEELFTESNSEVESEVYTEE 120
DB 61 tgftrsrealmaywekesqkllkerlgecgkvaedkeeseeeellftesenveesevytee 120
QY 121 EEEESQEEEDSDSEERTETETAKGNGVYVNSDNDKPKIFKSOIENINLTNGSN 180
DB 121 eeeesqeeeedseedeertletakgngvtyvnsdndskpkifksqienlnltngsn 180
QY 181 GRNTESPAIHPCGNPTVIEDALDKIKSNDPDTTEVNININENTITQTITRFAALKDNT 240
DB 181 grntespaahpcgnptviedaldkiksndpdttevnininentitqtitrfaealkdnt 240
QY 241 VVKFTSLANTADDSAAAMIAEMLKANEHTTNVNESNFTGKGLAIRMALQHTVYTE 300
DB 241 vvkftslantaddsaamaiaemlkanehttnvnesnftgkgilaalmalqhtvlyte 300
QY 301 LRFHNRHIMSGVEMETVTKLKENTTLRLGYHFEELPGPRMSMTSLTRMDQOROKRL 360
DB 301 lrfhngrhimgsvemeltvklkenttllrlgyhfeelpgprmsmtsltrmdqyrqkrl 360
QY 361 QEOKQOEGYDGGPNLRTKVMQGRTPSSSPYVSPRHSFWSFKLPKKQVTVSRPLSPVAT 420
DB 361 qeqkqgegydggpnlrktvwqgrtpssspyyvsprrhsfwsfplpkkqvtrsrplspvat 420
QY 421 LPPPPPPPPPPSSQRLPPPPPPPPPLPEKKLITRNIAEVITKQESQALQNGQKKK 480
DB 421 lppppppppppssqrlppppppppplpekklltrniaevitkqesaqalngqgkkk 480
QY 481 KGRKVKKOPNSILKEIKNSLSRVOEKMEKEDSSRSTPORSAHENLMEAIRGSSIKOLKRV 540
DB 481 kgkvvkkgpnslkellknsllsrvoekmedssrstrpqrshenlmeairgssikqlkry 540
```